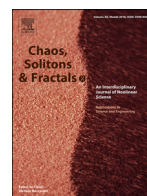




Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



# Evaluation and prediction of COVID-19 in India: A case study of worst hit states

Danish Rafiq\*, Suhail Ahmad Suhail, Mohammad Abid Bazaz

Department of Electrical Engineering, National Institute of Technology, Srinagar 190006, India

## ARTICLE INFO

### Article history:

Received 31 May 2020

Accepted 12 June 2020

Available online 19 June 2020

### Keywords:

SARS-CoV-2

Pandemic

System identification

## ABSTRACT

In this manuscript, system modeling and identification techniques are applied in developing a prognostic yet deterministic model to forecast the spread of COVID-19 in India. The model is verified with the historical data and a forecast of the spread for 30-days is presented in the 10 most affected states of India. The major results suggest that our model can very well capture the disease variations with high accuracy. The results also show a steep rise in the total cumulative cases and deaths in the coming weeks.

© 2020 Elsevier Ltd. All rights reserved.

## 1. Introduction

The advent and spread of 2019 novel coronavirus (SARS-CoV-2) has posed a global health crisis with a sharp rise in cases and deaths since its first detection in Wuhan, China, in December 2019. The infection causes illness ranging from common cold to extreme respiratory disease and death [1]. Currently, the prime epidemiological risk factor for 2019 novel coronavirus disease includes close contact with infected individuals with an incubation period of 2–14 days [2]. The case mortality rate is projected to range from 2 to 3% [3]. Various drugs are being assessed in line with previous researches into therapeutic treatments for SARS and MERS, however, there is no robust evidence for any significantly improved clinical outcome [4]. Apparent risk of acquiring the disease has led many governments to institute a variety of control procedures like quarantine, isolation and lock-down measures. Despite rigorous global containment measures, the frequency of the novel coronavirus disease continues to rise, with over 4.5 million confirmed cases and over 300,000 deaths worldwide as on 17<sup>th</sup> May, 2020 [5]. Although countries around the world have enhanced capacity building of the laboratory systems and response procedures, yet, there is a need for proper disease surveillance systems. Comprehending the initial transmission of the virus and analyzing the effectiveness of control measures are crucial in assessing the prospects for continued transmission in newer locations. This necessitates tracking the course of the pandemic to be able to foresee its emergence for a better response.

Prospective studies on modeling and forecasting of the epidemic have been carried out to provide analytical predictions on the size and end phase of the spread. Wu et al. [6] have used a susceptible exposed infectious recovered (SEIR) meta-population model to simulate the epidemic across all major cities in China. Early dynamics of transmission and control of COVID-19 within and outside Wuhan has also been studied using a stochastic transmission dynamic model [7]. Another study used the SEIR compartmental model to predict the feasibility for conducting the summer Olympics of 2020 in Japan [8]. Similarly, Abdullah et al. [9] presented a stochastic SIR model to predict the spread of COVID-19 in Kuwait. A classical SEIR type mathematical model is also presented by Mandal et al. [10] to study the qualitative dynamics of COVID-19 in India. Further work has been carried out by Ndairou et al. [11], with special focus on the transmissibility of super-spreader individuals in Wuhan, China.

Besides the above mentioned compartmental models, some other methods have been used to model and forecast the COVID-19 spread. For example, in Tomar and Gupta [12], a data-driven estimation method like long short-term memory (LSTM) is used for the prediction of total number of COVID-19 cases in India for a 30-days ahead prediction window. In addition to this, global epidemic and mobility model (GLEAM), an agent-based mechanistic model has also been used for daily forecasts of COVID-19 activity [13]. Harun, et al. [14] have used Box-Jenkins (ARIMA) and Brown/Holt linear exponential smoothing methods to estimate and forecast the number of COVID-19 cases in the G8 countries. Furthermore, Alqaness et al. [15] have incorporated a modified version of flower pollination algorithm (FPA) coupled with the salp swarm algorithm (SSA) to forecast the number of cases of COVID-19 for ten days in China.

\* Corresponding author.

E-mail address: [danish\\_pha2007@nitsri.net](mailto:danish_pha2007@nitsri.net) (D. Rafiq).

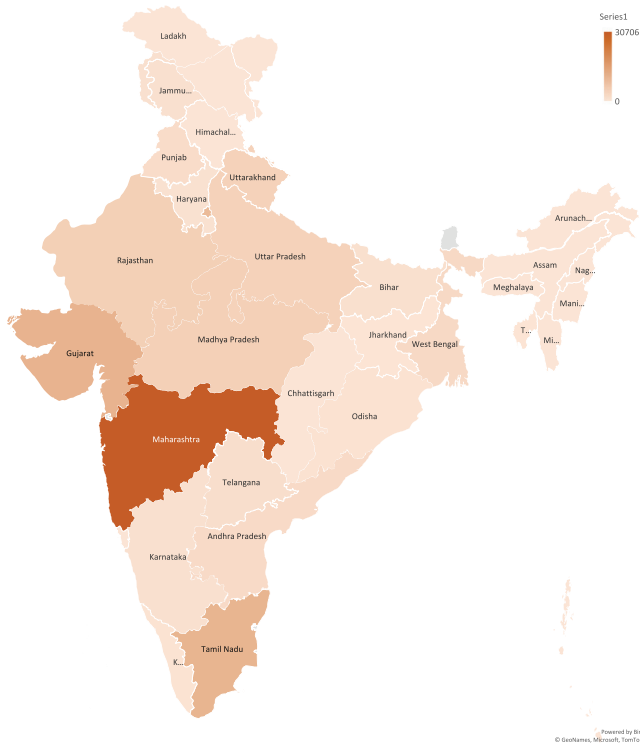


Fig. 1. Heat map of COVID-19 in Indian (as of 17 May 2020).

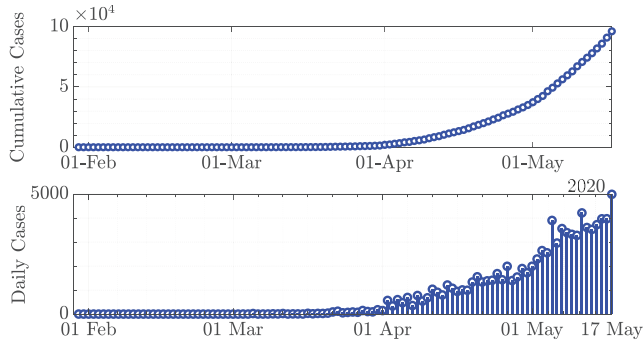


Fig. 2. (Top:) cumulative cases in India till 17 May 2020, (bottom:) daily new cases till 17 May 2020.

As on 17<sup>th</sup> May 2020, India has observed a total cases of 90,927 with 2872 deaths [16,17]. The very first case was reported on 30<sup>th</sup> January 2020, in a coastal state of Kerela (southern India) when a student returned from Wuhan, China. Subsequently, the number of positive cases in India rose rapidly due to the arrival of many passengers via airways [18]. An overview of the spread of COVID-19 in India is shown in Fig. 1. It can be easily seen that the virus has spread to entire country with the worst hit states being Maharashtra (30,706 cases), Gujarat (10,988), Tamil Nadu (10,588), Delhi (9333), Rajasthan (4960), and Madhya Pradesh (4789). Figs. 2 and 3 show the trend of rising new cases and deaths in India.

This manuscript demonstrates a control-theoretic, data-driven estimation technique to derive a time-series model from the historical data collected from [5,16] up-to 17<sup>th</sup> May 2020. The model is then used for the prediction of the total number of cases and deaths in most affected states of India for the next 30 days. The paper is sectioned as follows: Section 2 describes the system identification method employed. Section 3 presents the predicted cases

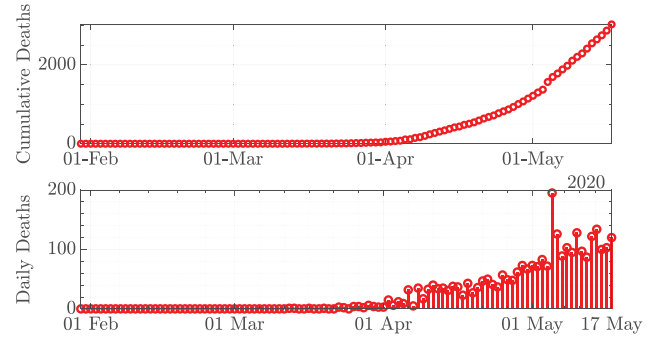


Fig. 3. (Top:) cumulative deaths in India till 17 May 2020, (bottom:) daily new deaths till 17 May 2020.

and deaths along-with some discussions. Finally, conclusions are presented in Section 4.

## 2. Data driven forecasting of COVID-19 in India

To estimate the spread of COVID-19 in India, we used a predictive error minimization (PEM) based system identification technique to identify a discrete-time, single-input, single-output (SISO) model [19–21]. Different models were identified for different states based on the data collected. The models were then verified on the testing data and upon validation, the models were used to predict the total number of cases and deaths for the next 30-days in the 10 worst hit states in India.

### 2.1. Model development

The discrete-time, identified model can be realized in the state-space from given as:

$$\begin{aligned} x(t + T_s) &= Ax(t) + Ku(t) \\ y(t) &= Cx(t) \end{aligned} \quad (1)$$

where the  $y(t)$  represents total number of cases or deaths of a particular area which is proportional to system state vector  $x(t) \in \mathbb{R}^n$ ,  $u(t)$  is the time series input and  $T_s$  is the sampling interval. Here, the unknowns to be identified are  $A \in \mathbb{R}^{n \times n}$ ,  $K \in \mathbb{R}^{n \times 1}$  and  $C \in \mathbb{R}^{1 \times n}$  which are in canonical form. Also,  $n$  is the dimension of the state-space model.

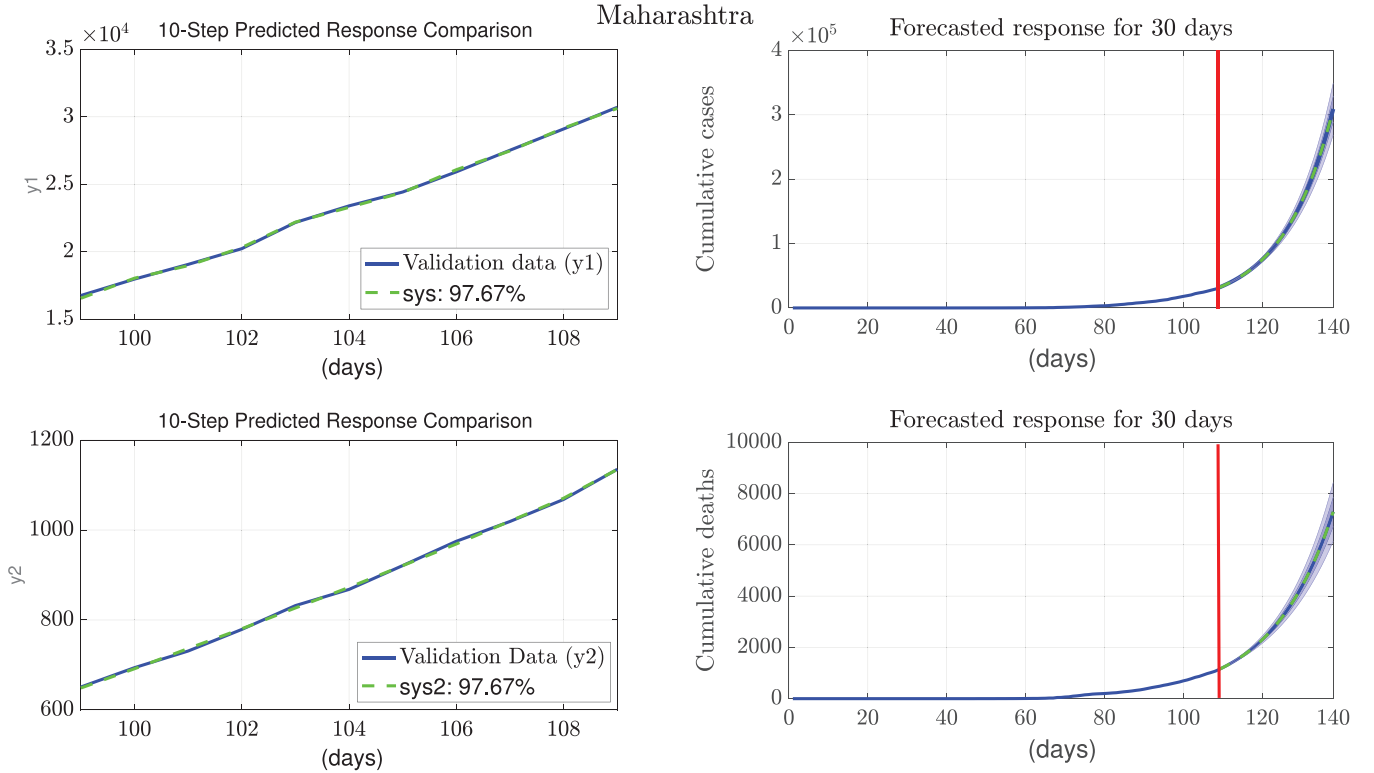
The identification problem can thus be posed as to selecting a model set  $\mathcal{M}(\theta)$  (indexed by a finite dimensional parameter vector  $\theta$ ) and evaluating a member from the set which best describes the recorded input-output relation according to a given criterion. One such criteria is given by Ljung [22] which is defined as :

$$v_N(\theta, z^N) = \frac{1}{N} \sum_{t=1}^N l(t, \theta, \epsilon(t, \theta)) \quad (2)$$

where  $\epsilon(t, \theta) = (y_0 - \hat{y}_0, \dots, y_N - \hat{y}_N)$  is referred as the prediction error,  $l(\cdot)$  is a scalar measure of fit,  $z(t) = [y^T(t), u^T(t)]$  and  $N$  is length of data-set. Typical choices of  $l(t, \theta, \epsilon)$  can be seen in Ljung [22].

The identified model thus minimizes the 1-step ahead prediction and the error  $\epsilon(t, \theta)$  between the measured  $y(t)$  and predicted values  $\hat{y}(t)$  is used to make the future prediction about the system. The prediction error identification estimate is thus given as:

$$\hat{\theta}_N = \arg \min_{\theta \in D_{\mathcal{M}}} v_N(\theta, z^N) \quad (3)$$



**Fig. 4.** (Top): 30-day prediction for number of cases in Maharashtra, (bottom): 30-day prediction for the number of deaths in Maharashtra. Red line shows the start of prediction window, dark blue:  $\pm 3$  std. deviation, light blue:  $\pm 5$  std. deviation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Here, we have taken:

$$\nu_N(\theta, z^N) = \frac{1}{N} \sum_{t=1}^N \frac{1}{2} \epsilon^2(t, \theta)$$

and the least-square problem has been solved iteratively via the *Levenberg-Marquardt* method [23–25].

The choice of model structure and its size is of crucial importance as it dictates the quality of long-term prediction and parameter estimation. The selection of model size  $n$  was made on the basis of the decay of the *Hankel* singular values of the system (1) [26,27].

### 3. Results and discussions

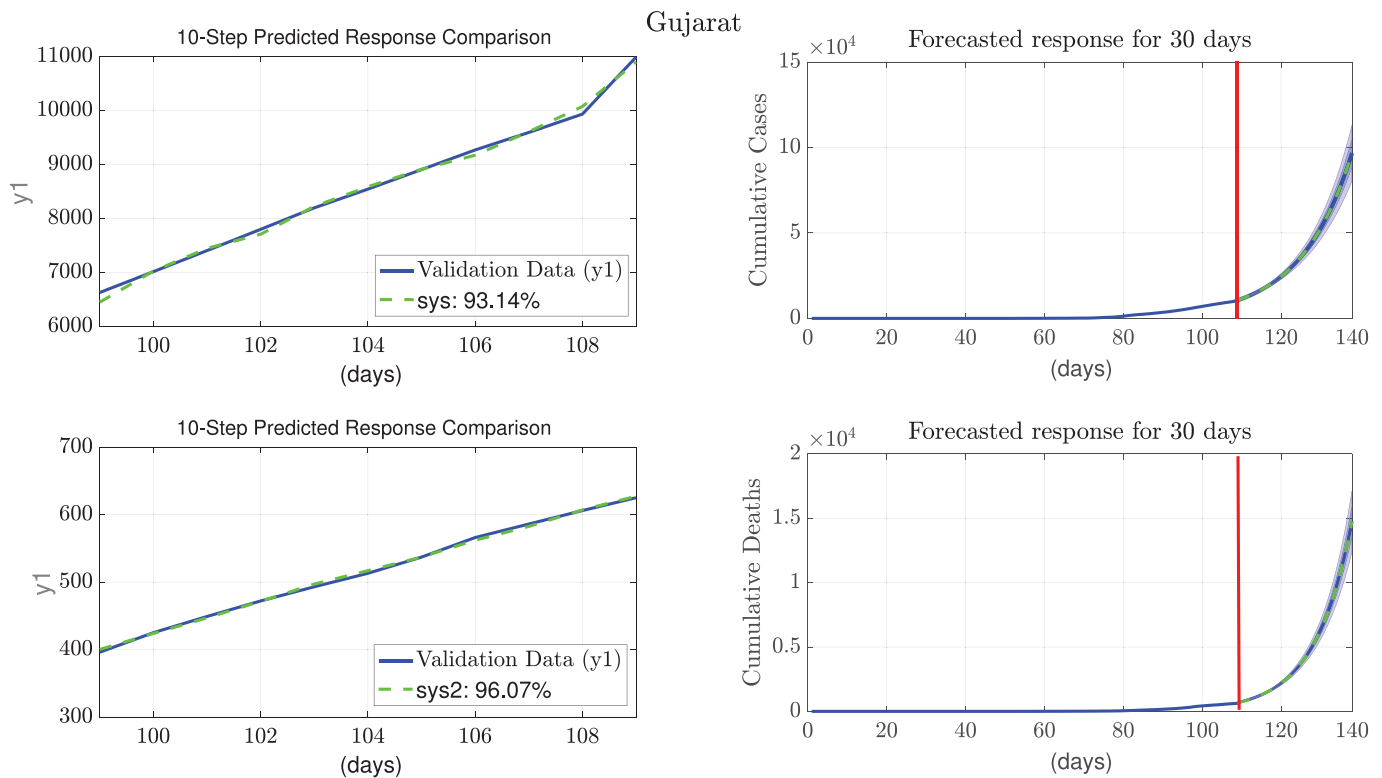
Figs. 4–13 show the dynamics of the forecasted response for the most infected states of India along-with a 10-step predicted response comparison with the validation data. Further results are presented in Table 1. As seen from Table 1, Maharashtra has recorded the highest number of COVID-19 cases accounting for 36% of the total country's caseload. It has also witnessed the sharpest rise in COVID-19 deaths with Mumbai being the epicenter of the pandemic in India. The constant influx of tourists, reliance on public transportation and population density have cumulatively made the metropolitan city hospitable for corona virus. Even though the state is conducting more tests, the violation of physical distancing rules by individuals particularly in containment zones result in the mixing of infected with healthy population. Moreover, unlike other red zones of Maharashtra, Mumbai faces shortage of ICU beds and dedicated COVID-19 hospitals. According to the prediction made herein, it would be inevitable that Mumbai and its suburbs would

continue to see an upsurge in the number of cases and deaths for at least up to 17<sup>th</sup> June 2020.

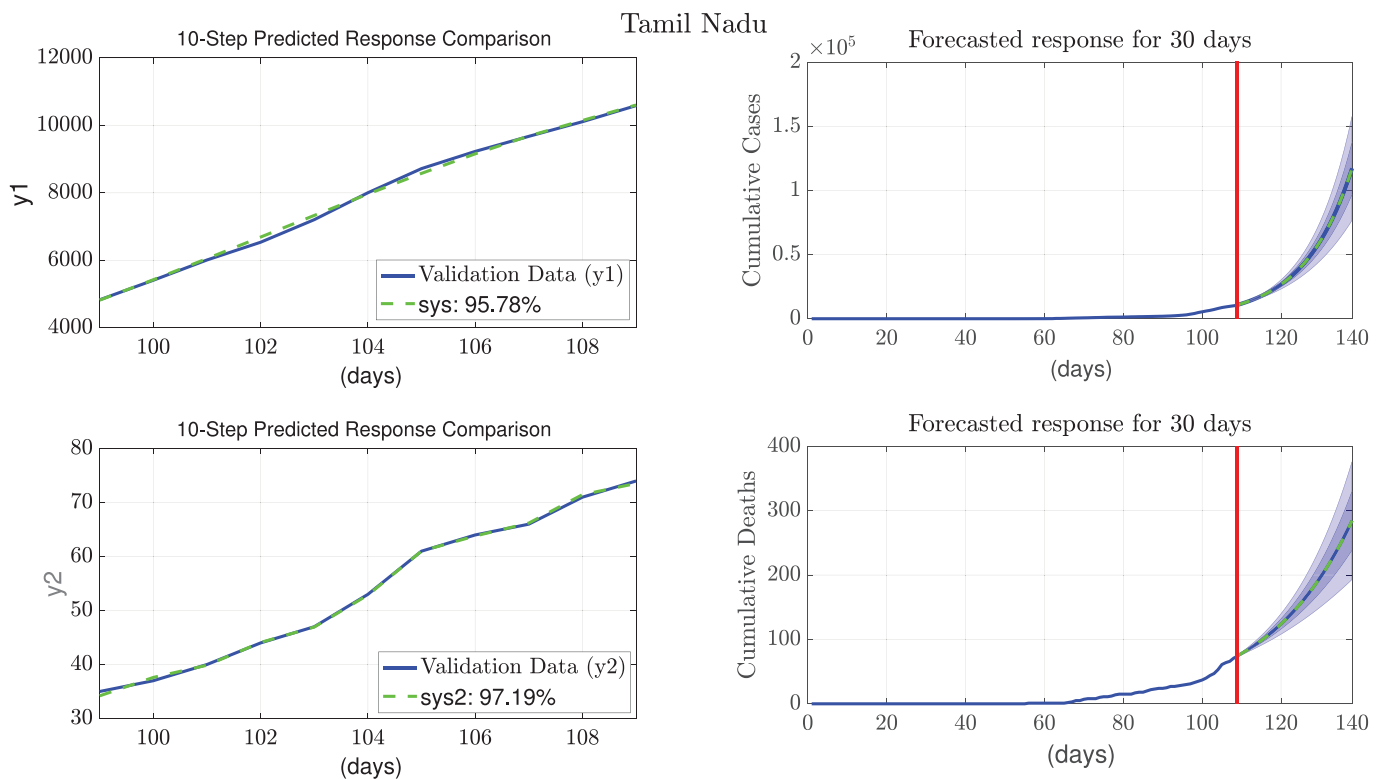
Gujarat has recorded the second highest COVID-19 mortality rate in the country in spite of reporting its first case as late as March 20. The COVID-19 mortality rate of Ahmedabad city is 6.8%, which is double the national average. Officials acknowledge that while Gujarat had its guard up sufficiently fast, there was a delay in testing. Even by mid of March, the daily average was as less as 15 tests per day, going up to 200/day by the end of March. According to the data driven identification scheme employed herein, the mortality rate in Gujarat may increase as high as 15.2% up to 17<sup>th</sup> June 2020.

Tamil Nadu, although being the third worst hit Indian state in terms of COVID-19 cases has witnessed the least number of mortalities with 1 among 143 positive cases succumbing to the disease (see Fig. 6). This is attributed to its credibility as a trusted medical center of the country. Chennai has the highest medical tourism in India with the state's average being above the national average in the health sector. This may be the reason that the predictable mortality rate of Tamil Nadu projected in this study is least among the rest of the states in consideration (see Table 1).

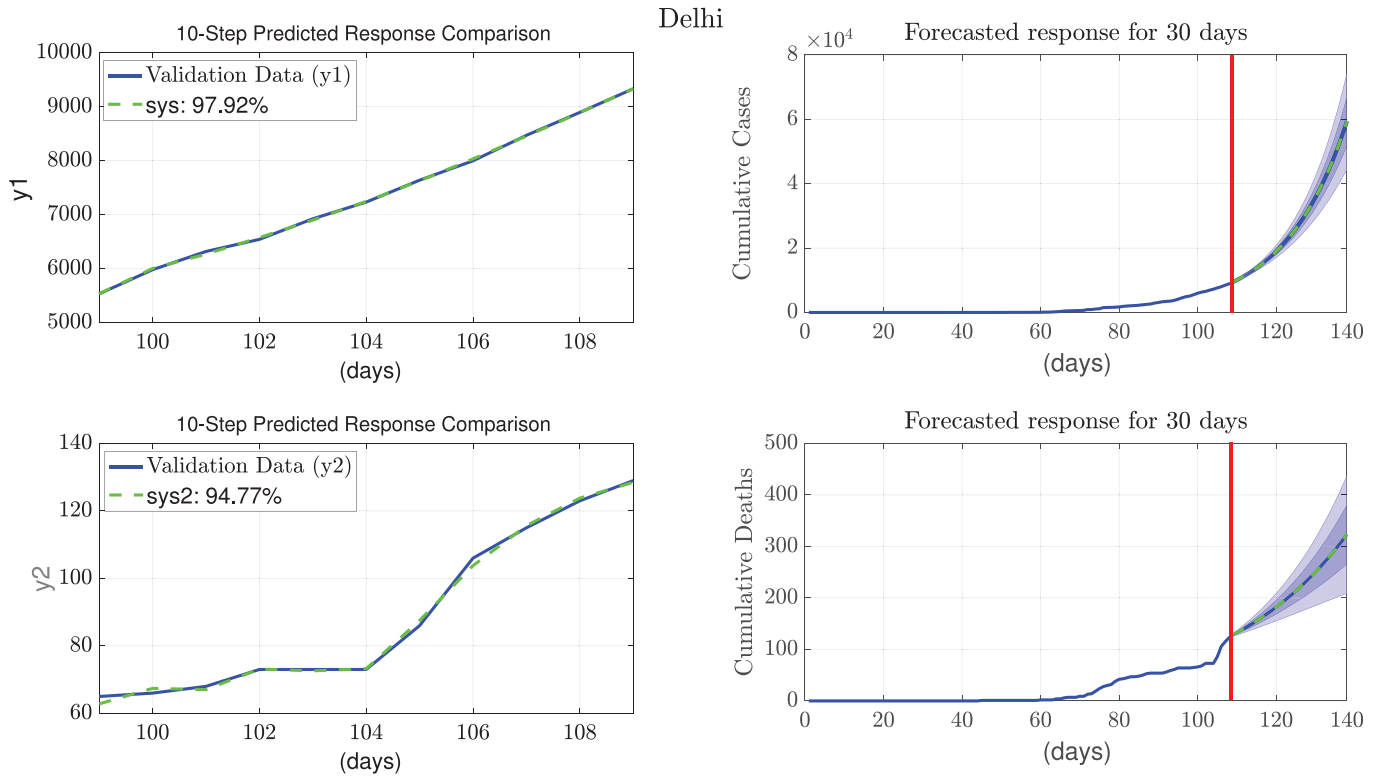
As per our prediction based on data up to 17<sup>th</sup> May 2020, Delhi along with other states would continue to see marginal surge in the number of COVID-19 cases owing to the relaxations in lock-down measures. The impact of removing the curbs will be more evident by the mid of June 2020. The under-funding of the health-care system, paucity of testing labs, violations of the lock-down protocols and inadequate quarantine facilities arranged by states and union territories are the biggest hurdles in combating the spread.



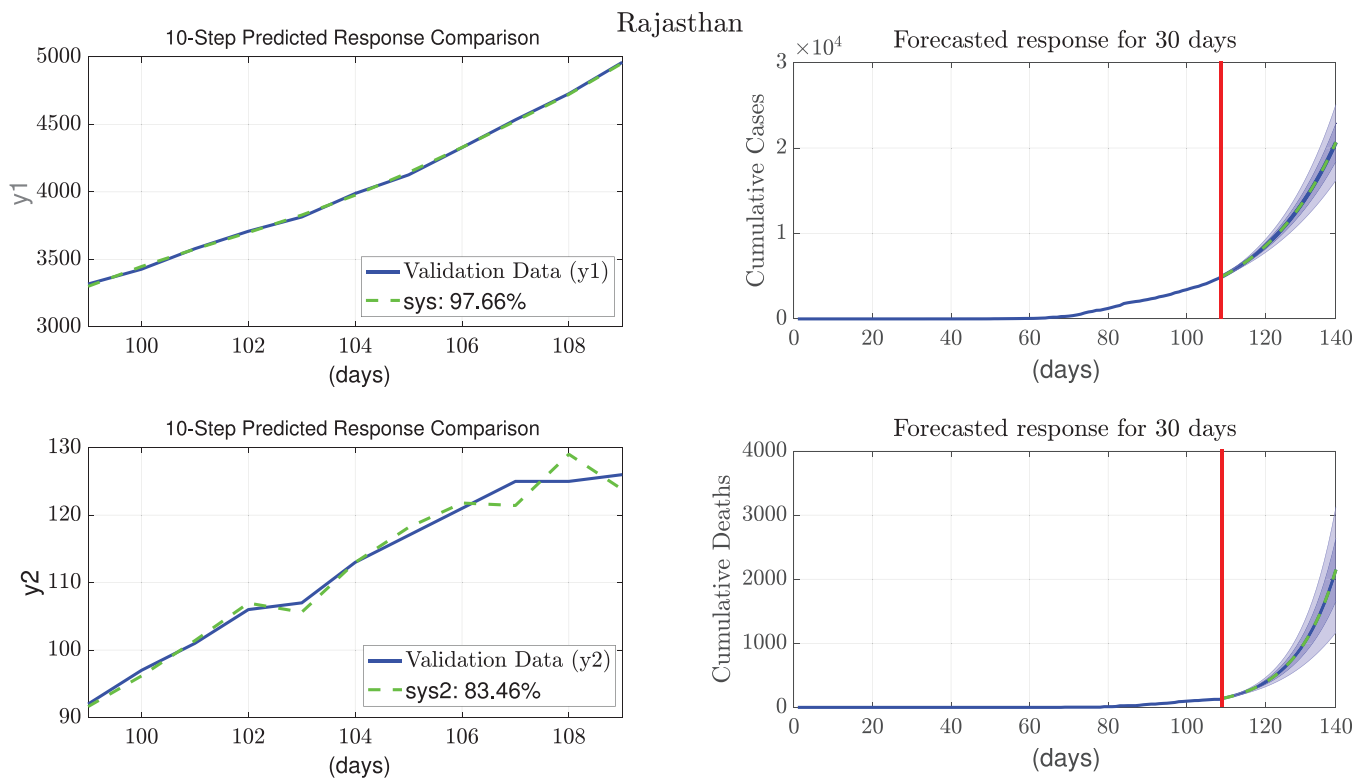
**Fig. 5.** (Top): 30-day prediction for number of cases in Gujarat, (bottom): 30-day prediction for the number of deaths in Gujarat. Red line shows the start of prediction window, dark blue:  $\pm 3$  std. deviation, light blue:  $\pm 5$  std. deviation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



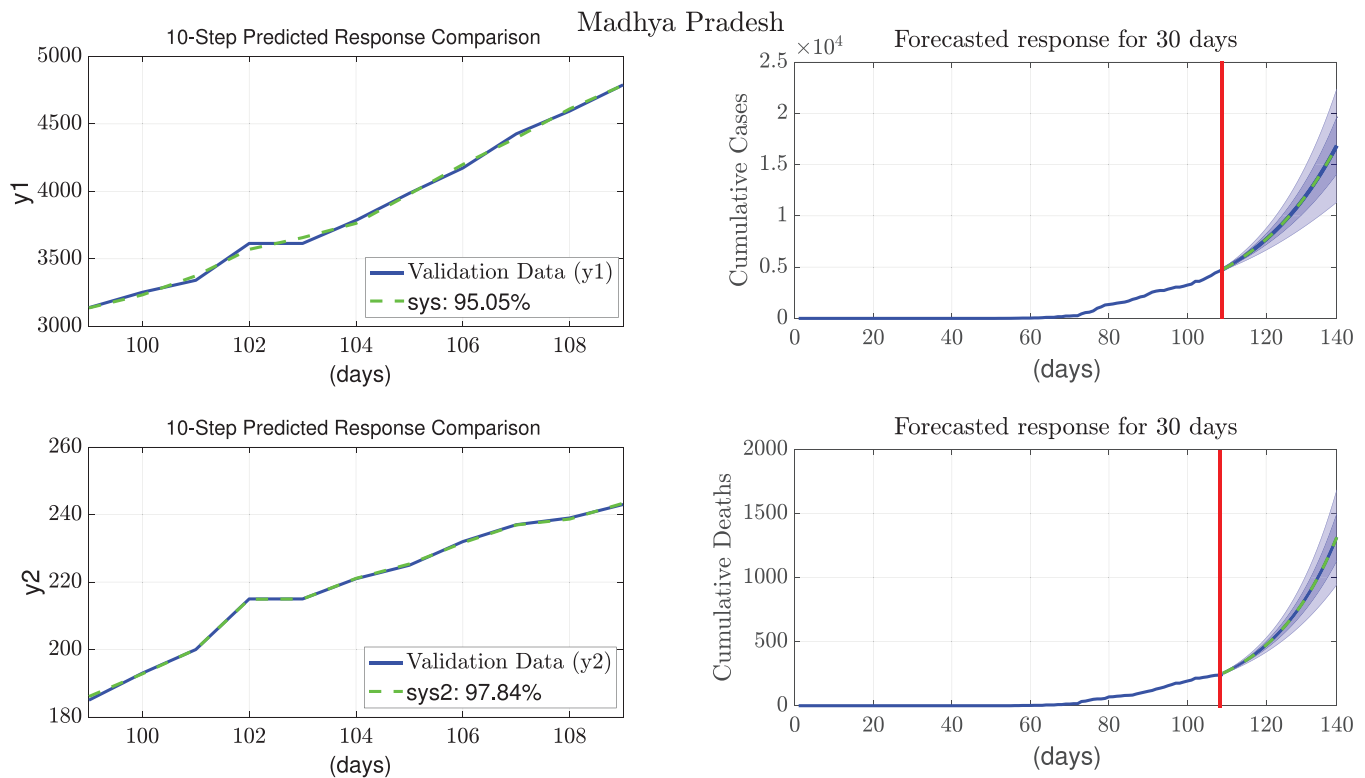
**Fig. 6.** (Top): 30-day prediction for number of cases in Tamil Nadu, (bottom): 30-day prediction for the number of deaths in Tamil Nadu. Red line shows the start of prediction window, dark blue:  $\pm 3$  std. deviation, light blue:  $\pm 5$  std. deviation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



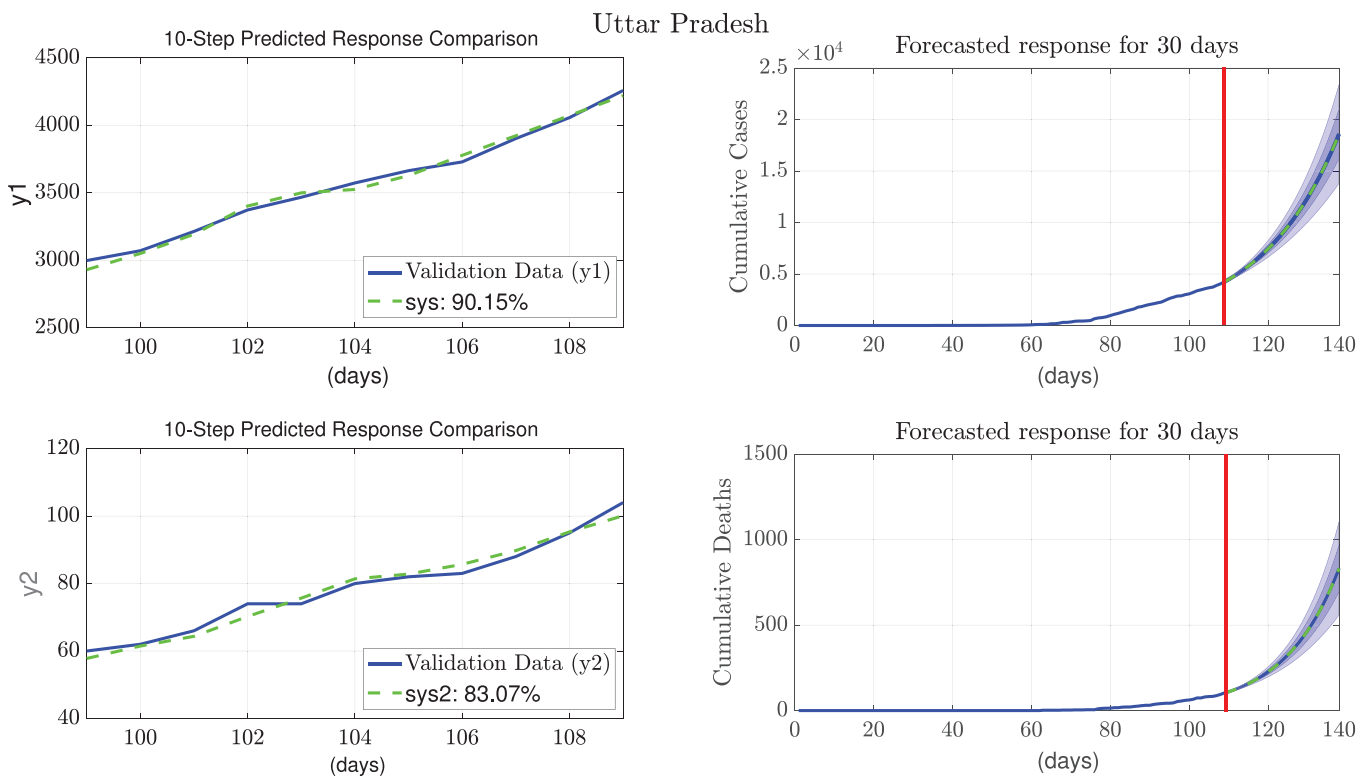
**Fig. 7.** (Top): 30-day prediction for number of cases in Delhi, (bottom): 30-day prediction for the number of deaths in Delhi. Red line shows the start of prediction window, dark blue:  $\pm 3$  std. deviation, light blue:  $\pm 5$  std. deviation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



**Fig. 8.** (Top): 30-day prediction for number of cases in Rajasthan, (bottom): 30-day prediction for the number of deaths in Rajasthan. Red line shows the start of prediction window, dark blue:  $\pm 3$  std. deviation, light blue:  $\pm 5$  std. deviation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

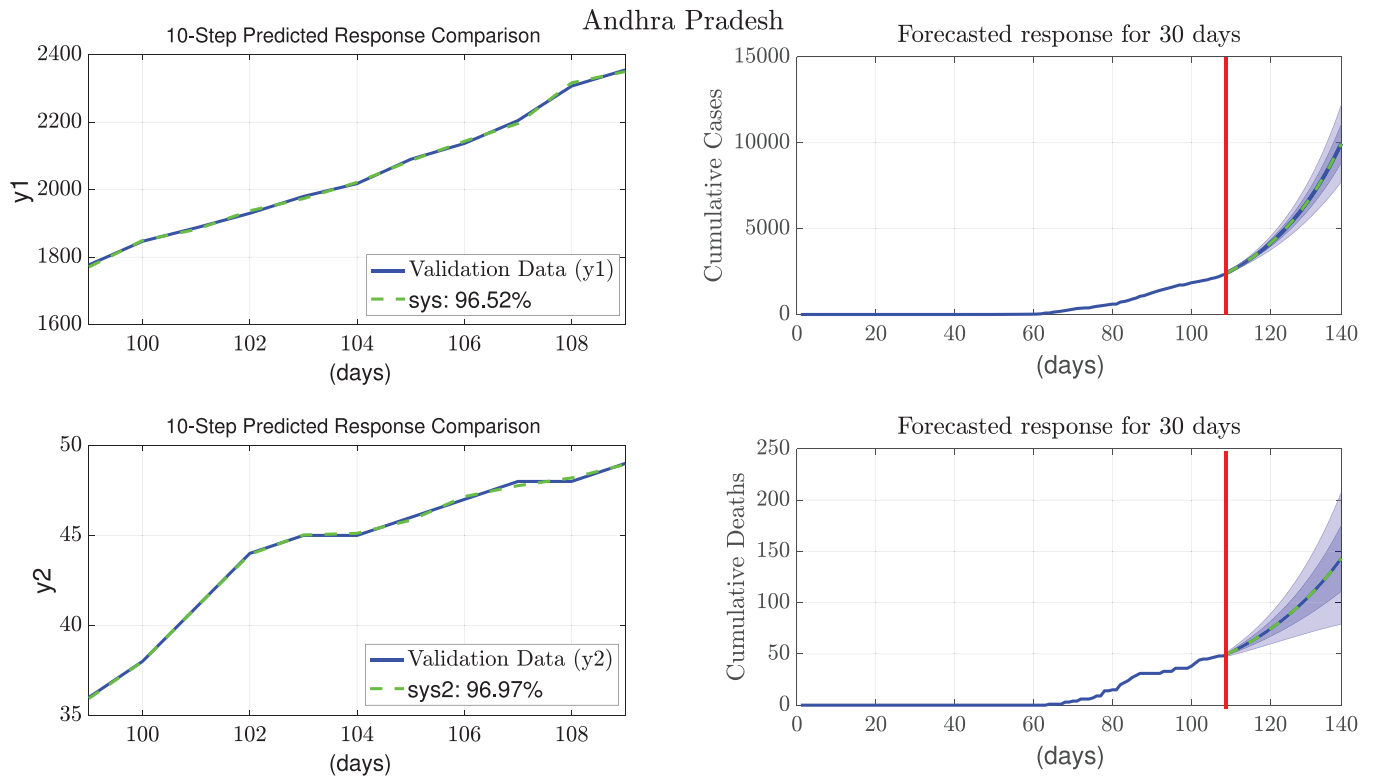


**Fig. 9.** (Top): 30-day prediction for number of cases in Madhya Pradesh, (bottom): 30-day prediction for the number of deaths in Madhya Pradesh. Red line shows the start of prediction window, dark blue:  $\pm 3$  std. deviation, light blue:  $\pm 5$  std. deviation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

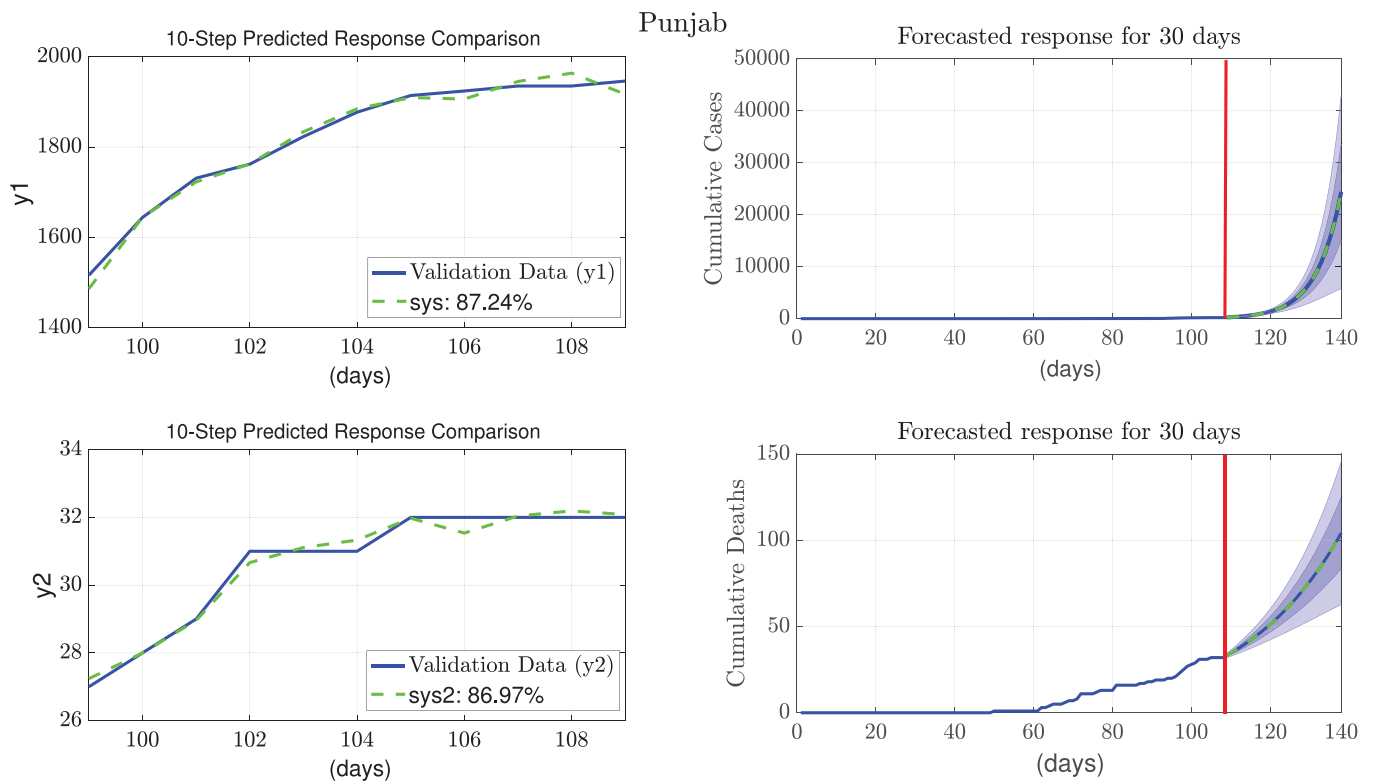


**Fig. 10.** (Top): 30-day prediction for number of cases in Uttar Pradesh, (bottom): 30-day prediction for the number of deaths in Uttar Pradesh. Red line shows the start of prediction window, dark blue:  $\pm 3$  std. deviation, light blue:  $\pm 5$  std. deviation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



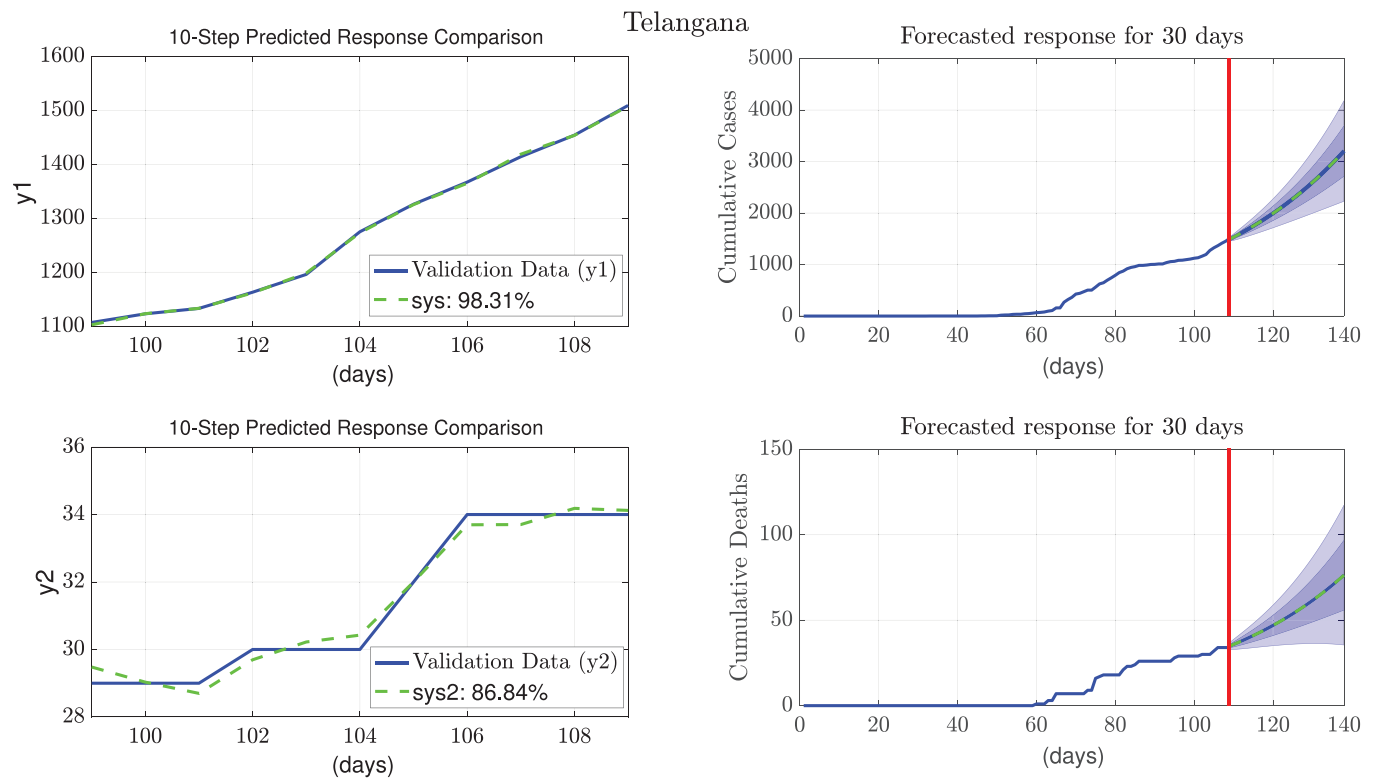


**Fig. 11.** (Top): 30-day prediction for number of cases in Andhra Pradesh, (bottom): 30-day prediction for the number of deaths in Andhra Pradesh. Red line shows the start of prediction window, dark blue:  $\pm 3$  std. deviation, light blue:  $\pm 5$  std. deviation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



**Fig. 12.** (Top): 30-day prediction for number of cases in Punjab, (bottom): 30-day prediction for the number of deaths in Punjab. Red line shows the start of prediction window, dark blue:  $\pm 3$  std. deviation, light blue:  $\pm 5$  std. deviation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)





**Fig. 13.** (Top): 30-day prediction for number of cases in Telangana, (bottom): 30-day prediction for the number of deaths in Telangana. Red line shows the start of prediction window, dark blue:  $\pm 3$  std. deviation, light blue:  $\pm 5$  std. deviation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

**Table 1**  
COVID-19 scenario in worst hit states of India upto 17 May 2020 along-with predicted values.

State	First case recorded (2020)	Total cases as on 17 Mar	Samples taken as on 17 Mar	Total deaths as on 17 Mar	Predicted cases upto 17 Jun	Predicted deaths upto 17 Jun	Predicted mortality rate upto 17 Jun
Maharashtra	10 Mar	30,760	2,474,040	1135	309,200	7289	2.35%
Gujarat	20 Mar	10,988	143,600	625	96,800	14,770	15.2%
Tamil Nadu	18 Mar	10,585	326,720	74	117,400	285	0.24%
Delhi	03 Mar	9333	135,791	129	59,430	323	0.54%
Rajasthan	03 Mar	4960	231,946	126	20,670	2150	10.4%
Madhya Pradesh	22 Mar	4789	103,898	243	16,860	1315	7.79%
Uttar Pradesh	04 Mar	4258	172,219	104	18,600	832	4.47%
Andhra Pradesh	19 Mar	2355	238,998	44	9959	143	1.43%
Punjab	18 Mar	1946	51,812	32	24,410	104	0.42%
Telangana	14 Mar	1509	13,750	34	3207	76	2.36%

#### 4. Conclusions

The study concerns the spread of COVID-19 in India. A control-theoretic approach is used to develop an epidemic model to simulate and predict the disease variations in 10 most affected states of India. Results depict a rapid increase in the number of cases in the coming days. However, it is pertinent to mention that the future estimation provided, is subject to certain system parameters and can vary based on the external inputs like lock-down measures, social-distancing, vaccine/drug development, rapid testing, etc. Information provided by our model could help establish a realistic assessment of the situation for the time-being and in the near future in order to apply the appropriate public health measures.

#### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Acknowledgments

The Doctoral fellowship of Author 1 and 2 from [Ministry of Human Resource Development \(MHRD/2017PHAELE006/009\)](#), New Delhi, India, is duly acknowledged. Author 1 would like to thank Asiya Batool for fruitful discussions.

## References

- [1] Gandhi R. T., Lynch J. B., del Rio C. Mild or moderate COVID-19. *N Engl J Med*. 2020. 10.1056/NEJMcp2009249.
- [2] Rafiq D., Batool A., Bazaz M. A. Three months of COVID-19: a systematic review and meta-analysis. *Rev Med Virol*. 2020. doi:10.1002/rmv.2113.
- [3] Mahase E. Coronavirus: COVID-19 has killed more people than SARS and MERS combined, despite lower case fatality rate. *BMJ* 2020;368.
- [4] Huang C. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 2020;395:497–506.
- [5] <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>.
- [6] Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-ncov outbreak originating in wuhan, china: a modelling study. *Lancet* 2020;395:689–97.
- [7] Kucharski A.J., et al. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *Lancet Infect Disease*. 2020. 10.1016/S1473-3099(20)30144-4.
- [8] Kuniya T. Prediction of the epidemic peak of coronavirus disease in japan. *J Clin Med* 2020;9:789.
- [9] Almeshal AM, Almazrouee AI, Alenizi MR, Alhajer SN. Forecasting the spread of COVID-19 in kuwait using compartmental and logistic regression models. *Appl Sci* 2020;10(10):3402. doi:10.3390/app10103402.
- [10] Mandal M, Jana S, Nandi SK, Khatua A, Adak S, Kar TK. A model based study on the dynamics of COVID-19: prediction and control. *Chaos Solitons Fractals* 2020. doi:10.1016/j.chaos.2020.109889.
- [11] Ndairou F, Area I, Nieto JJ, Torres DFM. Mathematical modeling of COVID-19 transmission dynamics with a case study of Wuhan. *Chaos Solitons Fractals* 2020. doi:10.1016/j.chaos.2020.109846.
- [12] Tomar A, Gupta N. Prediction for the spread of COVID-19 in India and effectiveness of preventive measures. *Sci Total Environ* 2020;728:138762. doi:10.1016/j.scitotenv.2020.138762.
- [13] Liu D., Clemente L., Poirier C., Ding X., Chinazzi M., Davis J.T., Vespignani A., Santillana M. A machine learning methodology for real-time forecasting of the 2019–2020 COVID-19 outbreak using internet searches, news alerts, and estimates from mechanistic models. 2020. D. <http://arxiv.org/abs/2004.04019>.
- [14] Yonar H. Modeling and forecasting for the number of cases of the COVID-19 pandemic with the curve estimation models, the Box-Jenkins and exponential smoothing methods. *Eurasian J Med Oncol* 2020;4(2):160–5. doi:10.14744/ejmo.2020.28273.
- [15] Al-qaness MA, Ewees AA, Fan H, Abd El Aziz M. Optimization method for forecasting confirmed cases of COVID-19 in China. *J Clin Med* 2020;9:674.
- [16] <https://www.covid19india.org/>.
- [17] <https://www.mohfw.gov.in/>.
- [18] <https://economictimes.indiatimes.com/news/politics-and-nation/15-286lakh-air-travellers-entered-india-in-2-months-cabinet287secretary/articleshow/74849839.cms?from=mdr>.
- [19] Ljung L. On the consistency of prediction error identification methods. *Math Sci Eng* 1976;126:121–64.
- [20] Aguirre LA, Billings SA. Dynamical effects of overparametrization in nonlinear models. *Phys D* 1995;80:26–40.
- [21] Aguirre LA, Billings SA. Improved structure selection for nonlinear models based on term clustering. *Int J Control* 1995;62:569–87.
- [22] Ljung L. System identification - theory for the user, appendix 4A. 2nd ed. Upper Saddle River, N.J.: PTR Prentice Hall; 1999.
- [23] Levenberg KA. Method for the solution of certain problems in least-squares. *Q Appl Math* 1944;2:164–8.
- [24] Marquardt D. An algorithm for least-squares estimation of nonlinear parameters. *SIAM J Appl Math* 1963;11:431–41.
- [25] Moré JJ. The Levenberg-Marquardt algorithm: implementation and theory, numerical analysis. In: Watson GA, editor. *Lecture notes in mathematics*, 630. Springer Verlag; 1977. p. 105–16.
- [26] [https://en.wikipedia.org/wiki/Hankel\\_singular\\_value](https://en.wikipedia.org/wiki/Hankel_singular_value).
- [27] Antoulas AC, Sorensen DC, Zhou Y. On the decay rate of Hankel singular values and related issues. *Syst Control Lett* 2002;5:323–42.